PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Oral switch versus standard intravenous antibiotic therapy in left-sided endocarditis due to susceptible staphylococci, streptococci or enterococci (RODEO): a protocol for two open-label randomised controlled trials
AUTHORS	Lemaignen, Adrien; BERNARD, Louis; Tattevin, P.; BRU, Jean-Pierre; Duval, Xavier; HOEN, Bruno; BRUNET HOUDARD, Solène; Mainardi, Jean-Luc; CAILLE, Agnes

VERSION 1 - REVIEW

REVIEWER	Goran Tešović
	University of Zagreb, School of Medicine
	University Hospital for Infectious Diseases
	Zagreb, CROATIA
REVIEW RETURNED	30-Sep-2019

GENERAL COMMENTS	Very good planned and organised multicenter national study covering topic of outstanding importance in clinical medicine. Since just one similarly designed study with significant number of participants has been published up to date, the present collaboration which tends to overcome the limitations of the previous study (POET) is expected to give important knowledge on the field of efficacious treatment of left-sided IE. Particular progres sin study design is visible through the idea of splitting the study ppulation into two arms („staphylococcal" and „streptococcal"). Distinguishing two different microbiologic entities with different natural biologic properties and clinical outcome would enable authors for better and more clear conclusions. I have no objections on the existing study design. Before the publishing the revision of some literature citatione which are not cited properly is needed.

REVIEWER	Marcos C Schechter
	Emory University School of Medicine, United States
REVIEW RETURNED	14-Oct-2019

GENERAL COMMENTS	Major comments: -The authors state the per protocol population will be defined prior to analysis. I do not see a definition of the PP population in the manuscript. I am not sure it is worth publishing the trial protocol

without a definition of the PP population.

-Patients will be followed for 6 months, but primary outcome is composite score of poor outcomes at 3 months. Suggest author explain why events occurring 3-6 following completion of antibiotic therapy are excluded from primary outcome analyses.

-Page 13, lines 50-54: "Symptomatic embolic events defined as secondary osteo-articular, splenic or brain localization after randomization". Understanding those are the most usual sites of metastatic disease, it is not clear to me why would embolic events to other organs be excluded from outcome. This is an ongoing trail and therefore protocol cannot be changed. Nonetheless, would be helpful for authors to explain rationale.

-Page 15. Secondary outcome, complains with oral therapy. Please clarify what is this "patient leaflet". Is it a validated tool? Who fills it, clinician or patient? By "return of treatment boxes", do authors mean pill count?

-Page 16, sample size justification. "taking into account the fact that we will only enrol patients who have a favourable outcome after the first two weeks of IE treatment" The protocol states will enroll patients after 10 days of IV antibiotics. Please clarify if 10 or 14 days.

-I do not have expertise review sample size justification. If this is a concern, another reviewer with this skillset will be needed.

Minor comments:

-Page 10, lines 52-57: "Both trials are based on the same protocol provided below, they are considered as distinct trials because sample size were calculated, so that each trial will be adequately powered." This sentence needs clarification.

-Page 13, lines 17-22: "For left-sided IE due to multi-susceptible Streptococcus sp., patients ≤ 70 kg receive amoxicillin 1500 mg three times daily and patients > 70 kg receive amoxicillin 2000 mg three times daily.". Suggest clarifying this will be the regimen for Enterococcal infections as well.

-Page 13, line 36: "Patients keep on IV antibiotic therapy for the remaining duration of treatment." I believe this is grammatically incorrect.

-Page 14. Definition of relapse. Authors could clearly state that microbiological relapse will not be assessed by genotypic tests.

-Page 14. Secondary outcome. Would state if control echocardiogram transthoracic or transesophageal.

-The authors could state how many patients recruited at this time

-Would add details on microbiological procedures. I assume all samples processed at local labs. Is antibiotic susceptibility testing standardized between centers?

-Should consider stratifying Staphylococcus trial between S aureus/S lugdunensis and other Staphylococci

REVIEWER	Irina Chis Ster
	St George's University of London
REVIEW RETURNED	20-Nov-2019

GENERAL COMMENTS	The clinical questions and the novelty of the research are beyond my expertise as a statistician.
	After careful considerations, I have an overall positive view on this manuscript describing the protocol of the two simultaneously conducted open-label randomized trials aiming to assess non-inferiority of oral switch as compared to entirely intravenous antibiotic therapy for the treatment of left-sided IE.
	I notice that the team has engaged with a professional statistician which is commendable.
	The sample size is correctly evaluated assuming data independency. I wonder whether the multi-centre aspect of the trial does not actually impose a hierarchical structure of the data, i.e. cluster structure and hence with consequences on the sample size evaluation.
	I would also like to see the sample size setting and calculations more clearly and the software used to produce it. I am not entirely sure what authors mean by "we assumed an expected failure rate of less than 10%" presumably for the control group. Also, I don't think that rate is the appropriate epidemiological measure in this stance; risk maybe?
	Please set out clearly (formally) the null and alternative hypotheses for these non-inferiority trials.
	I am not that keen on detailed analyses plans in the absence of any data - at this stage of the projects the authors need to ensure that they powered the studies adequately and the planned collection would allow incisive analyses. However, I would like to see mentioning potential sensitivity analyses and some broad plan to deal with missing data.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Goran Tešović
Institution and Country:
University of Zagreb, School of Medicine
University Hospital for Infectious Diseases
Zagreb, CROATIA
Please state any competing interests or state 'None declared': none declared

Please leave your comments for the authors below
Very good planned and organised multicenter national study covering topic of outstanding importance in clinical medicine. Since just one similarly designed study with significant number of participants has been published up to date, the present collaboration which tends to overcome the limitations of the previous study (POET) is expected to give important knowledge on the field of efficacious treatment of left-sided IE. Particular progres sin study design is visible through the idea of splitting the study ppulation into two arms („staphylococcal" and „streptococcal"). Distinguishing two different

microbiologic entities with different natural biologic properties and clinical outcome would enable authors for better and more clear conclusions.
I have no objections on the existing study design.
Before the publishing the revision of some literature citatione which are not cited properly is needed.

We thank the reviewer for his positive comments. We made some modifications in the references as requested: background sentence for references 7 to 12 (P.8) and 11 (P.12), adding of reference 5 (Martin-Carvajal et al) page 20, and actualization for reference 20 (P.20).

Reviewer: 2
Reviewer Name: Marcos C Schechter
Institution and Country: Emory University School of Medicine, United States
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
Major comments:

-The authors state the per protocol population will be defined prior to analysis. I do not see a definition of the PP population in the manuscript. I am not sure it is worth publishing the trial protocol without a definition of the PP population.
We are not able to anticipate every situation that could be considered as a major protocol violation and thus excluded from the per protocol analysis. Per protocol population will be defined during the blind review. We completed the Statistical Analyses paragraph as follows: "The PP population will exclude patients for whom there is a clear major protocol violation as defined during a blind review prior to any statistical analysis".

-Patients will be followed for 6 months, but primary outcome is composite score of poor outcomes at 3 months. Suggest author explain why events occurring 3-6 following completion of antibiotic therapy are excluded from primary outcome analyses.

We choose an evaluation of the primary outcome at 3 months after the end of the treatment for several reasons:
-        Previous studies suggest that most of poor outcomes (mainly death related to IE) occur in the first 3 months after diagnosis (Sunder et al, Plos One 2019)
-        A shorter duration for the evaluation of the primary outcome is supposed to decrease the risk of lost-to-follow-up
-        A shorter evaluation has no evidence-based reason to favor one group over the other in this context
Furthermore, the evaluation of a composite score of poor outcome at the end of follow-up is scheduled as a secondary objective.

-Page 13, lines 50-54: "Symptomatic embolic events defined as secondary osteo-articular, splenic or brain localization after randomization". Understanding those are the most usual sites of metastatic disease, it is not clear to me why would embolic events to other organs be excluded from outcome. This is an ongoing trail and therefore protocol cannot be changed. Nonetheless, would be helpful for authors to explain rationale.
We thank the reviewer for this comment. Indeed, we have specified the most usual sites of metastatic diseases but it is not an exhaustive list. The central point is the symptomatic characteristic of the embolic event: a silent embolism will not be considered as a failure, because in clinical practice, its diagnosis will be made in a fortuitous way, and therefore the timing of its apparition will not be exploitable.

On the contrary, a vascular embolism responsible for acute limb ischemia after randomisation will be considered as a failure for example.

We specified it in the manuscript: "Symptomatic embolic events defined as secondary osteo-articular, splenic, brain or other symptomatic localization after randomisation" (P.12).

-Page 15. Secondary outcome, complains with oral therapy. Please clarify what is this "patient leaflet". Is it a validated tool? Who fills it, clinician or patient? By "return of treatment boxes", do authors mean pill count?

There is no "gold-standard" for adherence evaluation, and literature suggests the use of several methods, including self-reported adherence (Anghel LA, Patient Prefer Adherence 2018). In order to evaluate patient's adherence to treatment, two methods are used in our study. The "patient leaflet" represents the self-reporting method. In practice, this leaflet is filled by the nurse in charge of the patient during hospitalization, and by the patient or his caregivers after returning home. This tool has not been specifically evaluated at our knowledge. The return of treatment boxes by the patient permits a pill count. Both methods are integrated in order to evaluate the adherence.

We specified these points in the manuscript: "The assessment of compliance with oral antibiotic treatment will be carried out at each visit during the treatment period by 2 combined methods: through a "patient leaflet" which will permit to note take/omissions of treatment, filled by the clinician during hospitalization, and by the patient or his caregivers after returning home; and through the return of the treatments' boxes to the pharmacy of the investigational site, thus allowing a pill count." (P.14)

-Page 16, sample size justification. "taking into account the fact that we will only enrol patients who have a favourable outcome after the first two weeks of IE treatment" The protocol states will enroll patients after 10 days of IV antibiotics. Please clarify if 10 or 14 days.

We thank the reviewer for this precision. As enounced in the "study intervention" section, patients may be randomized after 10 to 28 days of IV antibiotic treatment. Only patients with at least 14 days of remaining treatment can be included.

We corrected it in the "Sample size" section.

-I do not have expertise review sample size justification. If this is a concern, another reviewer with this skillset will be needed.

Minor comments:

-Page 10, lines 52-57: "Both trials are based on the same protocol provided below, they are considered as distinct trials because sample size were calculated, so that each trial will be adequately powered." This sentence needs clarification.

We clarified the sentence as follows: "Both trials are based on the same protocol provided below. Nevertheless, they are considered as two distinct trials, and sample sizes were calculated separately so that each trial has 80% power to show noninferiority of oral switch as compared to standard intravenous antibiotic therapy."

-Page 13, lines 17-22: "For left-sided IE due to multi-susceptible Streptococcus sp., patients ≤ 70 kg receive amoxicillin 1500 mg three times daily and patients > 70 kg receive amoxicillin 2000 mg three times daily.". Suggest clarifying this will be the regimen for Enterococcal infections as well.

This regimen is also applicated for patients with enterococcal infections. We specified it in the manuscript: "For left-sided IE due to multi-susceptible Streptococcus sp. or Enterococcus sp (i.e susceptible to amoxicillin with a minimal inhibitory concentration (MIC) ≤ 0.5mg/L), patients ≤ 70 kg receive amoxicillin 1500 mg three times daily and patients > 70 kg receive amoxicillin 2000 mg three times daily." (P.12).

-Page 13, line 36: "Patients keep on IV antibiotic therapy for the remaining duration of treatment." I believe this is grammatically incorrect.
We changed for : "patients continue IV antibiotic…"

-Page 14. Definition of relapse. Authors could clearly state that microbiological relapse will not be assessed by genotypic tests.
We specified it in the manuscript: "(i.e. same species, same antibiotic susceptibility profile, the realization of genotypic testing is not mandatory and left to the discretion of investigator)." (P.13).

-Page 14. Secondary outcome. Would state if control echocardiogram transthoracic or transesophageal.
The control should be made with the same method as that used for the diagnosis. This point is left to the discretion of the investigator.

-The authors could state how many patients recruited at this time
At the time of submission, 96 patients have been included in the RODEO 1 trial (staphylococci) and 190 in the RODEO 2 trial (streptococci/enterococci). We added this information in the manuscript at the end of the "Setting" section.

-Would add details on microbiological procedures. I assume all samples processed at local labs. Is antibiotic susceptibility testing standardized between centers?
All samples are processed at local labs and susceptibility testing comply with the EUCAST guidelines. In order not to overload the manuscript, we did not add this precision in the manuscript.

-Should consider stratifying Staphylococcus trial between S aureus/S lugdunensis and other Staphylococci
This point was not foreseen in the initial analysis plan and therefore cannot be modified a posteriori. This could be the subject of a post hoc sensitivity analysis.

Reviewer: 3
Reviewer Name: Irina Chis Ster
Institution and Country: St George's University of London
Please state any competing interests or state 'None declared': None declared.

Please leave your comments for the authors below

The clinical questions and the novelty of the research are beyond my expertise as a statistician. After careful considerations, I have an overall positive view on this manuscript describing the protocol of the two simultaneously conducted open-label randomized trials aiming to assess non-inferiority of oral switch as compared to entirely intravenous antibiotic therapy for the treatment of left-sided IE. I notice that the team has engaged with a professional statistician which is commendable. The sample size is correctly evaluated assuming data independency.

We thank the reviewer for these positive comments.

I wonder whether the multi-centre aspect of the trial does not actually impose a hierarchical structure of the data, i.e. cluster structure and hence with consequences on the sample size evaluation. Multi-centre randomized trials are indeed often stratified on centre, which then invites to perform an adjusted analysis (1). However, such a stratification is, to our knowledge, not taken into account in the sample size calculation. Taking it into account would lead to a smaller sample size (2) which may explain that in a conservative approach, people do not take it into account. Another issue is that to

take it into account supposes to a priori specify the center effect, which can be expressed as an intraclass correlation coefficient (2), such a piece of information is rarely available while planning a study. For the RODEO trial, randomization was not stratified on the center because of the high number of involved centers (n=46) and we stratified our randomization sequence on whether or not the patient underwent valvular surgery for the control of the current IE episode which is known to be associated with the patient's outcome. Stratification on center could have led to overstratification(3).Therefore there is no need to take it into account in sample size calculation.

1. Kahan BC. Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how? BMC Med Res Methodol. 2014 Feb 10;14:20.
2. Vierron E, Giraudeau B. Design effect in multicenter studies: gain or loss of power? BMC Med Res Methodol. 2009;9(39):39.
3. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. J Clin Epidemiol. 1999 Jan;52(1):19–26.

I would also like to see the sample size setting and calculations more clearly and the software used to produce it. I am not entirely sure what authors mean by "we assumed an expected failure rate of less than 10%" presumably for the control group. Also, I don't think that rate is the appropriate epidemiological measure in this stance; risk maybe?

We agree with the reviewer that "rate" was not appropriate. It was replaced by « proportion ». Same modification was made in the whole manuscript.

Please set out clearly (formally) the null and alternative hypotheses for these non-inferiority trials. Sample size calculations for this non-inferiority trial are based on a null hypothesis of H0: $\pi_2 - \pi_1 \geq$ delta (ie, inferior); where π_1 is the proportion of patients expected to experience failure in the intravenous group, π_2 is the proportion in the oral switch group, and the non-inferiority margin delta is 10%. The alternative hypothesis is $\pi_2 - \pi_1 <$ delta (ie, noninferior).

We specified it in the manuscript in the "Sample Size" subsection.

I am not that keen on detailed analyses plans in the absence of any data - at this stage of the projects the authors need to ensure that they powered the studies adequately and the planned collection would allow incisive analyses. However, I would like to see mentioning potential sensitivity analyses and some broad plan to deal with missing data.

Missing data will not be replaced except for the primary outcome on ITT population. Missing value will be considered an event whatever the randomised group. A sensitivity analysis will be performed excluding patients with missing primary outcome (complete-case analysis). Potential post-hoc sensitivity analyses will be performed.

We specified it in the manuscript: "Missing data will not be replaced except for the primary outcome on ITT population. Missing value will be considered a failure whatever the randomised group. A sensitivity analysis will be performed excluding patients with missing primary outcome (complete-case analysis). Potential post-hoc sensitivity analyses will be performed" (P.17).

VERSION 2 – REVIEW

REVIEWER	Marcos Schechter
	Emory University School of Medicine, United States
REVIEW RETURNED	06-Jan-2020

GENERAL COMMENTS	The authors have addressed most of my queries. I believe they mitigated as best as possible my concern with the per protocol population description. Given the importance of drug-susceptibility

testing (DSTs) protocol and reproducibility for their research, I would suggest clarifying where the DSTs will done and according to which standards. I am don't think this would add so much length to the paper, but will leave this decision to the authors and editors.

REVIEWER	Irina Chis Ster
	St George's University of London
REVIEW RETURNED	13-Jan-2020

| GENERAL COMMENTS | The clinical questions and the novelty of the research are beyond my expertise as a statistician. After careful considerations, I have an overall positive view on this manuscript describing the protocol of the two simultaneously conducted open-label randomized trials aiming to assess non-inferiority of oral switch as compared to entirely intravenous antibiotic therapy for the treatment of left-sided IE. I notice that the team has engaged with a professional statistician which is commendable. The sample size is correctly evaluated assuming data independency.
| | We thank the reviewer for these positive comments.
| | I wonder whether the multi-centre aspect of the trial does not actually impose a hierarchical structure of the data, i.e. cluster structure and hence with consequences on the sample size evaluation.
| | Multi-centre randomized trials are indeed often stratified on centre, which then invites to perform an adjusted analysis (1). However, such a stratification is, to our knowledge, not taken into account in the sample size calculation. Taking it into account would lead to a smaller sample size (2) which may explain that in a conservative approach, people do not take it into account. Another issue is that to take it into account supposes to a priori specify the center effect, which can be expressed as an intraclass correlation coefficient (2), such a piece of information is rarely available while planning a study. For the RODEO trial, randomization was not stratified on the center because of the high number of involved centers (n=46) and we stratified our randomization sequence on whether or not the patient underwent valvular surgery for the control of the current IE episode which is known to be associated with the patient's outcome. Stratification on center could have led to overstratification(3).Therefore there is no need to take it into account in sample size calculation.
| | 1. Kahan BC. Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how? BMC Med Res Methodol. 2014 Feb 10;14:20.
| | 2. Vierron E, Giraudeau B. Design effect in multicenter studies: gain or loss of power? BMC Med Res Methodol. 2009;9(39):39.
| | 3. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. J Clin Epidemiol. 1999 Jan;52(1):19–26.
| | New comments to the authors' response
| | I thank you the authors for the attempt to answering the question but it looks to me that they didn't understand it or I haven't been explicit enough. I was not asking about stratification (or adjustment) upon the centres - of course that would be inefficient. Stratification and clustering in the context of hierarchical data are two very different concepts. I was wondering about the clustering effect in the data given the considerable number if centres and patients within medical centres (46). The paper (1) the authors cite is a good start but my |

question was about the impact on the effect size the consideration of hierarchical structure of the data (patients within hospitals) would have. the conclusions the end of paper (1) does not make much sense to me.

"Fixed-effects, random-effects, or GEE with non-robust SEs should be used with a small number of centres. With a moderate or large number of centres, we recommend the use of either random-effects or GEE with a non-robust SE. "
I recommend the statistician to read on clustered randomised trials and I hope s/he will address my question. This is important as ignoring the hierarchical structure in the data may result in very small p-values as the potential variability between clusters remains unexplained.

I would also like to see the sample size setting and calculations more clearly and the software used to produce it. I am not entirely sure what authors mean by "we assumed an expected failure rate of less than 10%" presumably for the control group. Also, I don't think that rate is the appropriate epidemiological measure in this stance; risk maybe?
We agree with the reviewer that "rate" was not appropriate. It was replaced by « proportion ». Same modification was made in the whole manuscript.
Please set out clearly (formally) the null and alternative hypotheses for these non-inferiority trials. Sample size calculations for this non-inferiority trial are based on a null hypothesis of H0: $\pi_2 - \pi_1 \geq$ delta (ie, inferior); where π_1 is the proportion of patients expected to experience failure in the intravenous group, π_2 is the proportion in the oral switch group, and the non-inferiority margin delta is 10%. The alternative hypothesis is $\pi_2 - \pi_1 <$ delta (ie, noninferior). We specified it in the manuscript in the "Sample Size" subsection.
I am not that keen on detailed analyses plans in the absence of any data - at this stage of the projects the authors need to ensure that they powered the studies adequately and the planned collection would allow incisive analyses. However, I would like to see mentioning potential sensitivity analyses and some broad plan to deal with missing data.

Missing data will not be replaced except for the primary outcome on ITT population. Missing value will be considered an event whatever the randomised group. A sensitivity analysis will be performed excluding patients with missing primary outcome (complete-case analysis). Potential post-hoc sensitivity analyses will be performed.
We specified it in the manuscript: "Missing data will not be replaced except for the primary outcome on ITT population. Missing value will be considered a failure whatever the randomised group. A sensitivity analysis will be performed excluding patients with missing primary outcome (complete-case analysis). Potential post-hoc sensitivity analyses will be performed" (P.17).

New comments to the authors' response
I am not sure I understand how the authors have in mind addressing the missing data – accepting that their pattern, reasons, etc are impossible to envisage. But the authors should use an appropriate jargon for missing data situations such as complete data analysis and/or observed data analyses in connection to PP and ITT analyses. Concepts related to assumptions such as "missing (completely) at random", "missing not at random", etc should be mention rather than "missing data will not be replaced…". A

	statement such as "Missing value will be considered a failure whatever the randomised group" – is rather obscure. There is a great body of literature on missing data with particular reference to clinical trials – I strongly advice the authors to get a grip of the concepts associated with the missing data paradigm. One good place to start is https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3767219/. Sensitivity analyses assuming extreme scenarios should also be part of the analyses. Post -hoc analyses are again a different concept which is not necessarily related to missing data.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 2
Reviewer Name: Marcos Schechter
Institution and Country: Emory University School of Medicine, United States
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
The authors have addressed most of my queries. I believe they mitigated as best as possible my concern with the per protocol population description. Given the importance of drug-susceptibility testing (DSTs) protocol and reproducibility for their research, I would suggest clarifying where the DSTs will done and according to which standards. I am don't think this would add so much length to the paper, but will leave this decision to the authors and editors.

We thank the reviewer for his return. According to his demand, we added a short description of bacteriological procedures used during the trial:
"All microbiological wards are certified ISO15-189 and follow the current CASFM/EUCAST guidelines [1]. Drug-susceptibility testing follow the EUCAST disk diffusion method [2] and MIC are determined by broth microdilution or calibrated diffusion strips."

1   Société Française de Microbiologie. CASFM / EUCAST : Société Française de Microbiologie Ed ; 2019. 2019. https://www.sfm-microbiologie.org/wp-content/uploads/2019/05/CASFM2019_V2.0_MAI.pdf (accessed 5 Feb 2020).
2   Matuschek E, Brown DFJ, Kahlmeter G. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin Microbiol Infect 2014;20:O255-266. doi:10.1111/1469-0691.12373
3   Kahan BC, Morris TP. Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects? Stat Med 2013;32:1136–49. doi:10.1002/sim.5667

Reviewer: 3
Reviewer Name: Irina Chis Ster
Institution and Country: St George's University of London
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

New comments to the authors' response
I thank you the authors for the attempt to answering the question but it looks to me that they didn't understand it or I haven't been explicit enough. I was not asking about stratification (or adjustment) upon the centres - of course that would be inefficient. Stratification and clustering in the context of

hierarchical data are two very different concepts. I was wondering about the clustering effect in the data given the considerable number if centres and patients within medical centres (46). The paper (1) the authors cite is a good start but my question was about the impact on the effect size the consideration of hierarchical structure of the data (patients within hospitals) would have. the conclusions the end of paper (1) does not make much sense to me.

"Fixed-effects, random-effects, or GEE with non-robust SEs should be used with a small number of centres. With a moderate or large number of centres, we recommend the use of either random-effects or GEE with a non-robust SE. "

I recommend the statistician to read on clustered randomised trials and I hope s/he will address my question. This is important as ignoring the hierarchical structure in the data may result in very small p-values as the potential variability between clusters remains unexplained.

We agree with the reviewer that in a multicentre trial, patients in the same centre tend to have correlated outcomes and this clustering effect can be accounted for using hierarchical models. In the RODEO trial, we could use either a linear mixed model with an identity link function to estimate the risk difference or a logistic random effect model to estimate the odds ratio between groups for the primary outcome (both with random effects for centres). But ignoring the clustering effect in an individually randomized controlled trial does not impact the treatment effect estimate and may result in overestimated standard errors (SEs) and thus overestimated p-values [3]. Ignoring the clustering effect may result in underestimated standard errors (SEs) and thus too small p-values in a cluster randomized controlled trial in which all individuals within a centre receive the same treatment [4]. Ignoring the centre effect in the individually randomized RODEO trial is thus a conservative approach. Moreover, accounting for the centre effect has been shown to have an impact on the width of the 95% confidence interval when the intracluster correlation coefficient is quite large (over 0.05). In the RODEO trial, the ICC for the primary outcome is expected to be low (<0.01) because the ICC is an objective "outcome"-type variable [5].

Finally, as the RODEO trial is a noninferiority trial we chose to favor a primary conservative analysis regarding the noninferiority hypothesis (i.e. possibly too wide 95%CI).

For the above reasons, we added that a sensitivity analysis accounting for centre-effect will be performed but this analysis will not be the primary analysis of the trial just an exploratory analysis.

"To assess the impact of a potential centre-effect, a sensitivity analysis of the primary outcome will be performed with a random-centre-effect model."

4   Donner A. Design and analysis of cluster randomization trials in health research. London: : Arnold 2000.

5   Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. Clin Trials 2005;2:99–107.

New comments to the authors' response
I am not sure I understand how the authors have in mind addressing the missing data – accepting that their pattern, reasons, etc are impossible to envisage. But the authors should use an appropriate jargon for missing data situations such as complete data analysis and/or observed data analyses in connection to PP and ITT analyses.
Concepts related to assumptions such as "missing
(completely) at random", "missing not at random", etc should be mention rather than "missing data will not be replaced…". A statement such as "Missing value will be considered a failure whatever the randomised group" – is rather obscure. There is a great body of literature on missing data with particular reference to clinical trials – I strongly advice the authors to get a grip of the concepts associated with the missing data paradigm. One good place to start is
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3767219/.

Sensitivity analyses assuming extreme scenarios should also be part of the analyses. Post -hoc analyses are again a different concept which is not necessarily related to missing data.

We completed the manuscript to explain more unambiguously how we are planning to address the missing data and use the Rubin's missing-data-mechanisms terminology as advised by the reviewer. Management of missing data description is now as follows: "This analysis will be performed in both the ITT and PP populations. In the ITT analysis, missing primary outcome data will be handled by assuming that patients with missing data have treatment failure whatever the randomised group (worst case single imputation, assuming data are missing not at random). A sensitivity analysis will be performed excluding patients with missing primary outcome (complete-case analysis, assuming that data are missing completely at random)."
"No imputation of missing data will be performed for the secondary outcomes."

VERSION 3 - REVIEW

REVIEWER	Irina Chis Ster
	St George's University of London
REVIEW RETURNED	17-Mar-2020

GENERAL COMMENTS	I understand that this is an ongoing trial already, the recruitment has started. The authors addressed my questions and I understand that the ICC in developed countries can be very small. The fact that the trial is a non-inferiority one is irrelevant relative to the potentially hierarchical structure of that data. They accept that they take a conservative approach.To my view, the sample size should be calculated as for the worst case scenario to ensure the necessary number of participants for incisive analyses. Even in a small ICC , say 0.005 (the median reported across UK GP centres) the authors should investigate the loss of power. As it stands now, the trial is under-powered and the planned analyses would produce misleading results which - conservative is a light word.
	The authors make references to "rate" rather than risk. A rate involves a time unit, i.e. monthly, daily, etc. The authors should use the appropriate epidemiological jargon.
	Lastly but not least: in the light of recent developments regarding COVID19 - what are the difficulties the authors envisage in data collection and what are the mitigation measure for the relatively smooth running of the trial?

VERSION 3 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:
Reviewer: 3
Reviewer Name
Irina Chis Ster
Institution and Country
St George's University of London
Please state any competing interests or state 'None declared':
None declared
I understand that this is an ongoing trial already, the recruitment has started. The authors addressed my questions and I understand that the ICC in developed countries can be very small. The fact that

the trial is a non-inferiority one is irrelevant relative to the potentially hierarchical structure of that data. They accept that they take a conservative approach. To my view, the sample size should be calculated as for the worst case scenario to ensure the necessary number of participants for incisive analyses. Even in a small ICC, say 0.005 (the median reported across UK GP centres) the authors should investigate the loss of power. As it stands now, the trial is under-powered and the planned analyses would produce misleading results which - conservative is a light word.

We disagree with that point of view and already developed our rationale in a previous reviewing process. We allow ourselves to offer you this justification again:
We agree with the reviewer that in a multicentre trial, patients in the same centre tend to have correlated outcomes and this clustering effect can be accounted for using hierarchical models. In the RODEO trial, we could use either a linear mixed model with an identity link function to estimate the risk difference or a logistic random effect model to estimate the odds ratio between groups for the primary outcome (both with random effects for centres). But ignoring the clustering effect in an individually randomized controlled trial does not impact the treatment effect estimate and may result in overestimated standard errors (SEs) and thus overestimated p-values [3]. Ignoring the clustering effect may result in underestimated standard errors (SEs) and thus too small p-values in a cluster randomized controlled trial in which all individuals within a centre receive the same treatment [4]. Ignoring the centre effect in the individually randomized RODEO trial is thus a conservative approach. Moreover, accounting for the centre effect has been shown to have an impact on the width of the 95% confidence interval when the intracluster correlation coefficient is quite large (over 0.05). In the RODEO trial, the ICC for the primary outcome is expected to be low (<0.01) because the ICC is an objective "outcome"-type variable [5].
Finally, as the RODEO trial is a noninferiority trial we chose to favor a primary conservative analysis regarding the noninferiority hypothesis (i.e. possibly too wide 95%CI).
For the above reasons, we added that a sensitivity analysis accounting for centre-effect will be performed but this analysis will not be the primary analysis of the trial just an exploratory analysis.
"To assess the impact of a potential centre-effect, a sensitivity analysis of the primary outcome will be performed with a random-centre-effect model."

4   Donner A. Design and analysis of cluster randomization trials in health research. London: : Arnold 2000.
5   Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. Clin Trials 2005;2:99–107.

The authors make references to "rate" rather than risk. A rate involves a time unit, i.e. monthly, daily, etc. The authors should use the appropriate epidemiological jargon.

We agree with this comment and changed the word in the manuscript.

Lastly but not least: in the light of recent developments regarding COVID19 - what are the difficulties the authors envisage in data collection and what are the mitigation measure for the relatively smooth running of the trial?

We thank the reviewer for this topical comment and added a comment in the manuscript in the "Setting" sub-section:
"During the COVID-19 crisis, the maintenance of new inclusions was left to the discretion of the Research Department of the participating centers from March 17 to May 11, 2020. However, the follow-up visits for the patients already included were maintained as planned, in teleconsultation if necessary."